# 83+ DATASETS

## RELEASED BY

Google

@learn.machinelearning

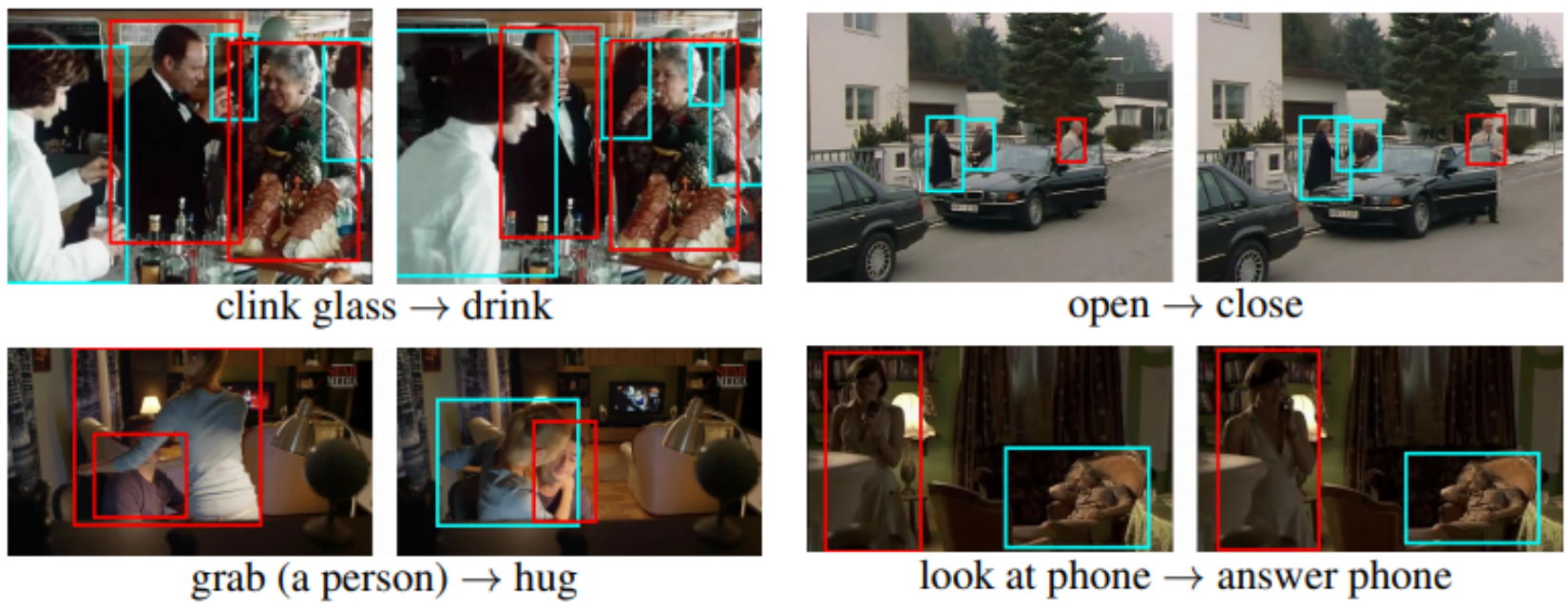
# 1. **AudioSet**

Among the popular audio datasets, AudioSet is a large-scale dataset of manually annotated audio events. It includes an expanding ontology of 632 audio event classes, and a collection of 20,84,320 human-labelled 10-second sound clips drawn from YouTube videos.

source - AIM

@learn.machinelearning

# 2. AVA Dataset




AVA is a video dataset of spatio-temporally localised Atomic Visual Actions (AVA) that provides audiovisual annotations of video to improve and understand human activity. The dataset annotates 80 atomic visual actions in 430 15-minute movie clips, where actions are localised in space and time. AVA dataset is a collection of 1.62 million action labels with multiple labels per human occurring frequently.

# 3. Cartoon Set

Cartoon Set is a collection of random, 2D cartoon avatar images where the cartoons vary in 10 artwork categories, four colour categories and four proportion categories, with a total of approximately 1,013 possible combinations. The cartoons in this dataset helped develop the technology behind the personalised stickers in Google Allo.

source - AIM



# 4. Coached Conversational Preference Elicitation

This dataset consists of 502 English dialogues with 12,000 annotated utterances between a user and an assistant discussing movie preferences in natural language. The dataset has been gathered using a Wizard-of-Oz methodology between two paid crowd-workers, where one worker plays the role of an ‘assistant’, while the other plays the role of a ‘user’.

## 5. DiscoFuse

DiscoFuse is a large-scale dataset for Discourse-Based Sentence Fusion (DiscoFuse) that includes approximately 60 million sentence fusion examples. Sentence fusion is the task of joining several independent sentences into a single coherent text.

source - AIM



## 6. Google’s Conceptual Captions

Google‘s Conceptual Captions dataset consists of approximately 3.3 million images annotated with captions. In contrast with the curated style of other image caption annotations, Conceptual Caption images and their raw descriptions are harvested from the web, and therefore represent a wider variety of styles.

# 7. Grasping Dataset

The Grasping Dataset contains roughly 8,00,000 plus grasp attempts over two months, using between 6 and 14 robotic manipulators at any given time, with differences in camera placement and hardware.

source - AIM



# 8. HDR+ Burst Photography Dataset

This dataset consists of 3,640 bursts that are made up of 28,461 images in total and organised into subfolders, including the results of the image processing pipeline. Each burst consists of the raw burst input in DNG format.

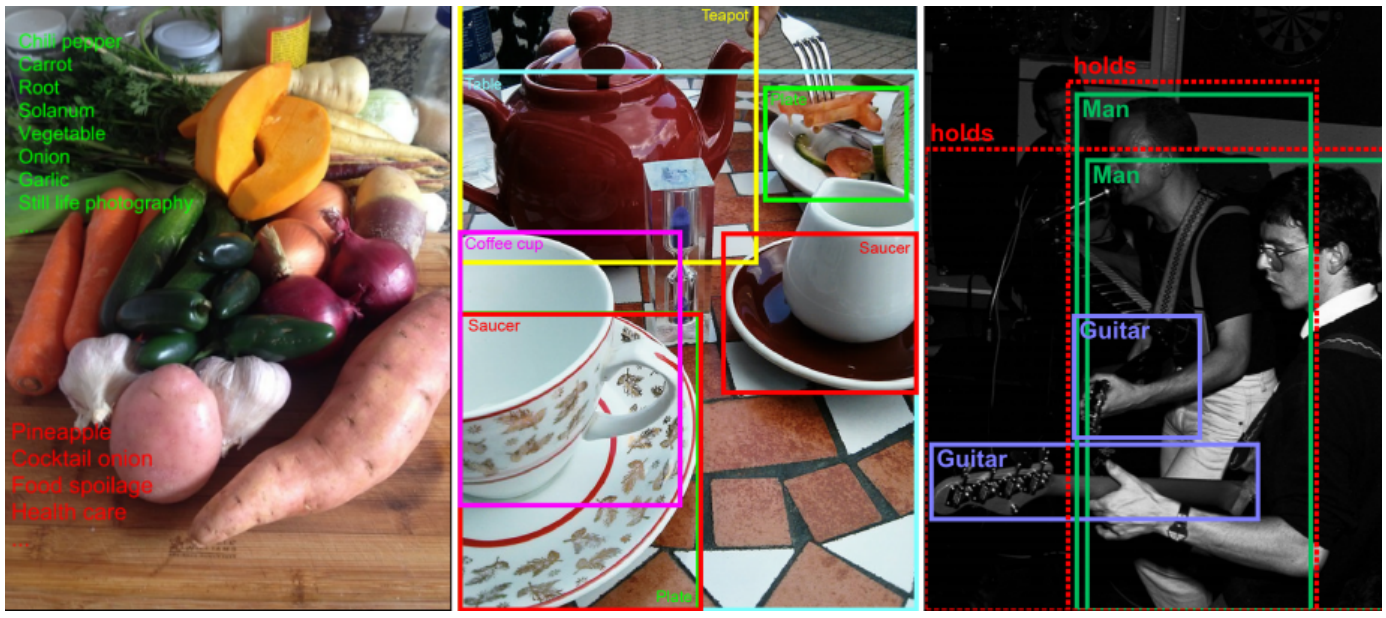
# 9. Noun Verb

This dataset contains naturally-occurring 30,000 English sentences that feature non-trivial noun-verb ambiguity. The dataset contains sentences in CoNLL format, and each sentence has a single token that has been manually annotated as either VERB or NON-VERB.

source - AIM

@learn.machinelearning

# 10. Open Images Dataset V6




The Open Images Dataset V6 is one of the popular datasets released by Google. It includes approximately 9 million images annotated with image-level labels, object bounding boxes, object segmentation masks, visual relationships, and localised narratives. The dataset contains 16 million bounding boxes for 600 object classes on 1.9 million images. This makes it the largest existing dataset with object location annotations.

# 11. RealEstate10K

RealEstate10K is a large dataset of camera poses corresponding to 10 million frames derived from about 80,000 video clips, gathered

from about 10,000 YouTube videos. For each clip, the poses form a trajectory where each pose specifies the camera position and orientation along the trajectory. These poses are derived by running SLAM and bundle adjustment algorithms on a large set of videos.



# 12. Taskmaster-1

The Taskmaster-1 dataset consists of 13,215 task-based dialogues in English, including 5,507 spoken and 7,708 written dialogues created with two distinct methods. In this dataset, each conversation falls into one of 6 domains. These are – ordering pizza, creating appointments for an auto repair, ride service set up, ordering movie tickets, ordering coffee drinks, and making reservations in restaurants.

# 13. The Quick, Draw! Dataset

The Quick Draw Dataset includes 50 million drawings across 345 categories that are contributed by players of the game Quick, Draw! The drawings were captured as timestamped vectors, which are tagged with metadata, including what the player was asked to draw and the location of the player.

source - AIM

@learn.machinelearning

# 14. The MAESTRO Dataset

The MIDI and Audio Edited for Synchronous Tracks and Organisation – or MAESTRO – dataset is a collection of over 200 hours of virtuosic piano performances, captured with fine alignment (~3 ms) between note labels and audio waveforms.

# 15. Taskmaster-2

The Taskmaster-2 dataset consists of 17,289 dialogues in seven domains: restaurants (3,276), food ordering (1,050), movies (3,047), hotels (2,355), flights (2,481), music (1,602), and sports (3,478). All dialogues in this dataset were collected using the same Wizard of Oz (WOz) system used in Taskmaster-1, where crowdsourced workers playing the “user” interacted with human operators playing the “digital assistant” using a web-based interface.



# 16. Youtube-8M Segments Dataset

The YouTube-8M Segments dataset is an extension of the YouTube-8M dataset with human-verified segment annotations. It is a collection of human-verified labels on about 2,37,000 segments on 1,000 classes from the validation set of the YouTube-8M dataset, where each video will again come with time-localised frame-level features, so that classifier predictions can be made at segment-level granularity.

# DATASET LINKS IN BIO

# Thank You.

Like, Comment, Share and Save it for Later

**Happy Machine Learning**

@learn.machinelearning