

Open Images Extended - Crowdsourced

Open Images Extended - Crowdsourced intends to capture global representation. This dataset comprises over 478,000 images and associated labels from otherwise under-represented populations. It can be used with Open Images V4.

PUBLISHER(S)

Google LLC

INDUSTRY TYPE

Corporate - Tech

KEY APPLICATION

Machine Learning, Object Recognition

INTENDED USE CASE(S)

- Identify objects or context of photos visually (e.g., through Lens or Camera)
- Find objects, plants, animals, etc. through search in Photos or Image Search

PRIMARY DATA TYPE

Image Data

DATASET CHARACTERISTICS

(All numbers are approximate)

Total Instances	478k+
Total Classes	6k+
Total Labels	1.27m+
Algorithmically Generated Labels	1.11m+
User Contributed Labels	505k+
Human Verified Labels	All labels verified

NATURE OF CONTENT

Labeled images of objects (household goods, commercial products), vehicles, plants, animals and people (faces blurred).

DATASET FUNCTION(S)

Training
Testing

EXCLUDED DATA

All [EXIF](#) data including location has been removed

PRIVACY [PII](#) associated with human subjects removed

LICENSE TYPE(S)

CC-BY-4.0

LAST UPDATED	Oct 2018
VERSION	1.0
STATUS	Actively Maintained

SUMMARY OF LICENSE PERMISSIONS ([CC-BY-4.0](#))

- You are free to share and adapt
- Attribution required
- You cannot apply any additional restrictions

ACCESS COST Open Access

DATA COLLECTION METHOD(S)

Crowdsourced

DATA SOURCE(S)

- Contributions by global users of the [Crowdsourcing](#) app
- Vendor data collection efforts

DATA SELECTION

All images are opted-in for open-sourcing by Crowdsourcing app contributors

SAMPLING METHOD(S)

Unsampled

GEOGRAPHIC DISTRIBUTION

83%	India
2%	Vietnam
2%	Brazil
1%	Israel
1%	Nigeria
1%	Thailand
1%	Colombia
1%	UAE
8%	Others (each less than 1%)

FILTERING CRITERIA

- PII: Name tags, Unblurred faces, etc.
- Inappropriate Content
- Unusable Imagery

LABELING METHOD(S)

Human Labels
Algorithmic Labels

LABEL TYPE(S)

Human Labels	Free-form text labels
Algorithmic Labels	Additional labels

LABEL SOURCE(S)

Human Labels	Image owners
Algorithmic Labels	Google's internal image annotation algorithm

LABELING PROCEDURE - HUMAN

Free-form labels are provided by users of the Crowdsourcing app. The user who has taken the picture provides the label.

LABELING PROCEDURE - ALGORITHMIC

Labels are resolved against known entity names from [Knowledge Graph](#). Additional labels are added based on Google's internal image annotation system.

VALIDATION METHOD(S)

Human Validated

VALIDATION TASK(S)

- Human validators verify labels
- Human validators flag PII
- Human validators filter data

VALIDATOR DESCRIPTION(S)

Compensated workers based out of India

VALIDATION POLICY SUMMARY

Algorithmic and user contributed labels are verified by human validators based out of India. There is a known overlap in algorithmic and user contributed labels. Validators flag any PII content.