

# Open Images Extended - More Inclusively Annotated People (MIAP)

[Dataset Download](#) [Related Publication](#)

This dataset was created for fairness research and fairness evaluations in person detection. This dataset contains 100,000 images sampled from Open Images V6 with additional annotations added. Annotations include the image coordinates of bounding boxes for each visible person. Each box is annotated with attributes for perceived gender presentation and age range presentation. It can be used in conjunction with Open Images V6.

## Authorship

PUBLISHER(S)

Google LLC

INDUSTRY TYPE

Corporate - Tech

DATASET AUTHORS

Candice Schumann, Google, 2021  
Susanna Ricco, Google, 2021  
Utsav Prabhu, Google, 2021  
Vittorio Ferrari, Google, 2021  
Caroline Pantofaru, Google, 2021

PUBLISHER(S)

Google LLC

FUNDING TYPE

Private Funding

DATASET CONTACT

[open-images-extended@google.com](mailto:open-images-extended@google.com)

## Motivations

DATASET PURPOSE(S)

Research Purposes

Machine Learning

Training, testing, and validation

KEY APPLICATION(S)

Machine Learning

Object Recognition

Machine Learning Fairness

PROBLEM SPACE

This dataset was created for fairness research and fairness evaluation with respect to person detection.

[See accompanying article](#)

PRIMARY MOTIVATION(S)

- Provide more complete ground-truth for bounding boxes around people.
- Provide a standard fairness evaluation set for the broader fairness community.

INTENDED AND/OR SUITABLE USE CASE(S)

- **ML Model Evaluation for:** person detection, fairness evaluation
- **ML Model Training for:** person detection, Object detection

Also:

- **Person detection:** Without specifying gender or age presentations
- **Fairness evaluations:** Over gender and age presentations
- **Fairness research:** Without building gender presentation or age classifiers

## Use of Dataset

SAFETY OF USE

Conditional Use

There are some known unsafe applications.

UNSAFE APPLICATION(S)



Gender classification

Age classification

UNSAFE USE CASE(S)

This dataset should not be used to create gender or age classifiers. The intention of perceived gender and age labels is to capture gender and age presentation as assessed by a third party based on visual cues alone, rather than an individual's self-identified gender or actual age.

CONJUNCTIONAL USE

Safe to use with other datasets

KNOWN CONJUNCTIONAL DATASET(S)

- The data in this dataset can be combined with [Open Images V6](#)

KNOWN CONJUNCTIONAL USES

Analyzing bounding box annotations not annotated under the Open Images V6 procedure.

METHOD

Object Detection

SUMMARY

A person object detector can be trained using the Object Detection API in Tensorflow.

KNOWN CAVEATS

If this dataset is used in conjunction with the original Open Images dataset, negative examples of people should only be pulled from images with an explicit negative person image level label. The dataset

does not contain any examples not annotated as containing at least one person by the original Open Images annotation procedure.

METHOD

Fairness Evaluation

SUMMARY

Fairness evaluations can be run over the splits of gender presentation and age presentation.

KNOWN CAVEATS

There still exists a gender presentation skew towards unknown and predominantly masculine, as well as an age presentation range skew towards middle.