Using Existing Datasets for **TinyML**

Don't *collect* from scratch

Data collection is **difficult**!

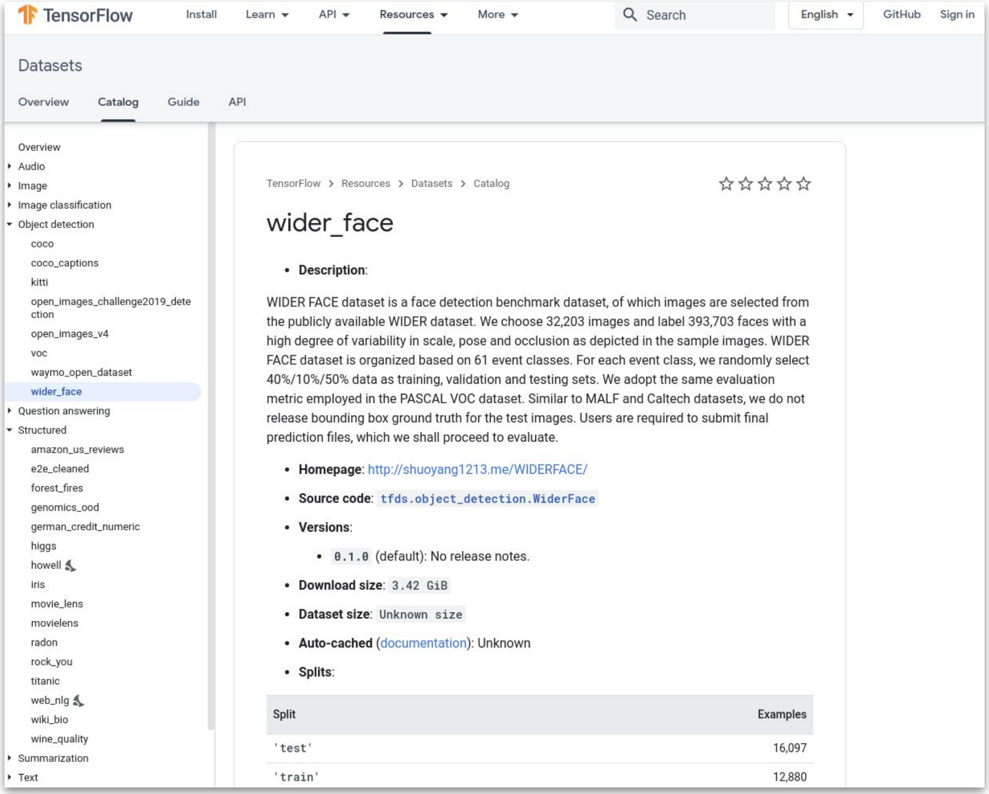
- Can we ***reuse*** existing data?

What's available?

What's missing?

TensorFlow
Datasets Catalog

Audio
Image
Image Classification
Object Detection
Question Answering
Structured
Summarization
Text
Translate
Video



wider_face

- **Description**:

WIDER FACE dataset is a face detection benchmark dataset, of which images are selected from the publicly available WIDER dataset. We choose 32,203 images and label 393,703 faces with a high degree of variability in scale, pose and occlusion as depicted in the sample images. WIDER FACE dataset is organized based on 61 event classes. For each event class, we randomly select 40%/10%/50% data as training, validation and testing sets. We adopt the same evaluation metric employed in the PASCAL VOC dataset. Similar to MALF and Caltech datasets, we do not release bounding box ground truth for the test images. Users are required to submit final prediction files, which we shall proceed to evaluate.

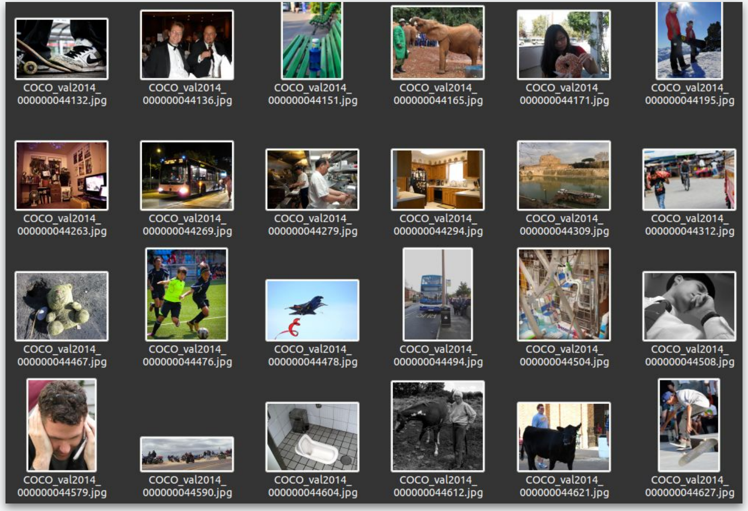
- **Homepage**: http://shuoyang1213.me/WIDERFACE/
- **Source code**: `tfds.object_detection.WiderFace`
- **Versions**:
  - `0.1.0` (default): No release notes.
- **Download size**: `3.42 GiB`
- **Dataset size**: `Unknown size`
- **Auto-cached** (documentation): Unknown
- **Splits**:

Split	Examples
'test'	16,097
'train'	12,880

TinyML
Person Detection

- **Visual Wake Words**: a new dataset built from Common Objects in Context (**COCO**)
    - ***people*** v. ***no people***

Repurposing existing datasets for **TinyML** tasks is a powerful concept


COCO_val2014_ 000000044132.jpg
COCO_val2014_ 000000044136.jpg
COCO_val2014_ 000000044151.jpg
COCO_val2014_ 000000044165.jpg
COCO_val2014_ 000000044171.jpg
COCO_val2014_ 000000044195.jpg
COCO_val2014_ 000000044263.jpg
COCO_val2014_ 000000044269.jpg
COCO_val2014_ 000000044279.jpg
COCO_val2014_ 000000044294.jpg
COCO_val2014_ 000000044309.jpg
COCO_val2014_ 000000044312.jpg
COCO_val2014_ 000000044467.jpg
COCO_val2014_ 000000044476.jpg
COCO_val2014_ 000000044478.jpg
COCO_val2014_ 000000044494.jpg
COCO_val2014_ 000000044504.jpg
COCO_val2014_ 000000044508.jpg
COCO_val2014_ 000000044579.jpg
COCO_val2014_ 000000044590.jpg
COCO_val2014_ 000000044604.jpg
COCO_val2014_ 000000044612.jpg
COCO_val2014_ 000000044621.jpg
COCO_val2014_ 000000044627.jpg

Don’t ***learn*** from scratch

- **Transfer** learning
- **Pretrained** models: your “AI Data Labeling Assistant”
- **Generate** your own data
  - Simulations
  - ML models