

The IntentQA dataset is contained in the file `pascal_gt.json`. The format is as follows:

```
{Img id: Points}
```

```
Points:
{String: [int]}
```

```
Point string: "x_y"
```

Each image is associated with points of the format "x\_y" (e.g. "186\_329"). Each point is associated with a list of labels, one per worker that labeled the point. Labels are in the range [1, 40]. There are 20 object classes in Pascal VOC, and for each class a point can be either a 'whole object' ([1, 20]) or 'part' ([21, 40]) point. The image IDs are from the Pascal VOC Dataset. The object classes not present in the IntentQA dataset are simply not present (for example, the labels 20/40, corresponding to 'tv monitor', are not present.)

`pascal_gt.json` consists of all 49,623 points collected. See `pascal_gt_clean.json` for the 48,143 points actually used during training.

The VQA-converted dataset is also available under `intentqa_vqa.json`. The format is as follows:

```
{Img id: [QA pair]}
```

```
QA Pair {
  Img_id: string
  Question: string
  Ans: string
  Point: Point
}
```

```
Point {
  x: int
  y: int
  ans: string
}
```

The image IDs are from Pascal VOC.