

The dataset can be found in `localqa_dataset.json`. The organization of the dataset is as follows:

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

```
QA Pair {  
  Img_id: int  
  Question: string  
  All_ans: [string]  
  All_objs: [Annotation]  
  Points: [Point]  
}
```

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

```
Annotation {  
  Names: [String]  
  Attributes: [String]  
  w: int  
  h: int  
  x: int  
  y: int  
  ...  
}
```

Note that a single QA pair consists of a single question (e.g. “What color is this shirt?”). The field `All_ans` indicates all answers to the question in the image; the field `All_objs` indicates the objects corresponding to each answer, and `Points` indicates the point associated with each answer (the center of the object’s bounding box). `All_objs` and `Points` contain the answers in the same order; each QA Pair is then “flattened” to ask a separate question for each of the objects in `All_objs`. Image IDs correspond to IDs in Visual Genome.

Image splits for train/val/test-dev/test-final can be found in the JSON files.