The datasets can be found in spatial\_dataset.json and verbal\_dataset.json. The organization of the spatial dataset is as follows:

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

QA Pair {
  Img_id: int
  Question: string
  Answer: string
  Point: Point
  }

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

The verbal dataset is similarly organized, except for no Point field.