|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| |  | | --- | | **Feature** | | |  | | --- | | **DAQUAR** | | |  | | --- | | **VQA V2.0** | |
| |  | | --- | | **Size** | | **1,449 Images** With **12,468 Question-Answer Pairs** | **204,721 Images** With Over **1.1 Million Question-Answer Pairs** |
| **Size On Disk** | **~ 500 MB** | **~ 20 GB** |
| **Sources Of Images** | Indoor Scenes | COCO Dataset (Both Indoor And Outdoor Scenes) |
| **Question Types** | Object Based Simple Questions | Diverse Object Based, Reasoning,” Yes/No,” “What/Why/When” |
| **Answer Type** | Short One-Word Answers | Varies In Single Word Phrases And Sentences |
| **Task Focus** | Object Recognition And Counting In Indoor Scenes | Broader, Including Reasoning, Relationships, Actions |
| |  | | --- | | **Evaluation Metric** |  |  | | --- | |  | | Accuracy-based | More complex accuracy, considering multiple correct answers |