|  |  |
| --- | --- |
| Datasets | Links |
| COCO-Karpathy | <https://storage.googleapis.com/sfr-vision-language-research/datasets/coco_karpathy_train.json> (lables)  <http://images.cocodataset.org/zips/train2014.zip> (images) |
| COCO VQA | <https://storage.googleapis.com/sfr-vision-language-research/LAVIS/datasets/vqav2/vqa_train.json> (vqa\_train)  https://storage.googleapis.com/sfr-vision-language-research/LAVIS/datasets/vqav2/vqa\_val.json (vaq\_val) |
| Visual Genome | <https://cs.stanford.edu/people/rak248/VG_100K_2/images.zip> (images\_part1)  <https://cs.stanford.edu/people/rak248/VG_100K_2/images2.zip> (images\_part2)  <https://homes.cs.washington.edu/~ranjay/visualgenome/data/dataset/image_data.json.zip> (labels) |
| TextCaps | <https://cs.stanford.edu/people/rak248/VG_100K_2/images.zip> (images)  <https://dl.fbaipublicfiles.com/textvqa/data/textcaps/TextCaps_0.1_train.json> (labels) |
| RefCOCO | <https://bvisionweb1.cs.unc.edu/licheng/referit/data/refcoco.zip> (labels) |
| RefCOCO+ | <https://bvisionweb1.cs.unc.edu/licheng/referit/data/refcoco+.zip> (labels) |
| RefCOCOg | <https://bvisionweb1.cs.unc.edu/licheng/referit/data/refcocog.zip> (labels) |
| OKVQA | https://storage.googleapis.com/sfr-vision-language-research/LAVIS/datasets/okvqa/okvqa\_train.json (labels) |
| AOK-VQA | https://storage.googleapis.com/sfr-vision-language-research/LAVIS/datasets/aokvqa/aokvqa\_v1p0\_train.json (labels) |
| OCR-VQA | https://drive.google.com/drive/folders/1\_GYPY5UkUy7HIcR0zq3ZCFgeZN7BAfm\_?usp=sharing (labels) |
| GQA | <https://downloads.cs.stanford.edu/nlp/data/gqa/images.zip> (images)  https://storage.googleapis.com/sfr-vision-language-research/LAVIS/datasets/gqa/train\_balanced\_questions.json (labels) |
| Filtered flickr-30k | <https://drive.google.com/drive/folders/19c_ggBI77AvdtYlPbuI0ZpnPz73T5teX?usp=sharing> (labels) |
| Multi-task conversation | <https://drive.google.com/file/d/11HHqB2c29hbSk-WLxdta-nG8UCUrcCN1/view?usp=sharing> (labels) |
| Filtered unnatural instruction | <https://drive.google.com/file/d/1lXNnBcb5WU-sc8Fe2T2N8J0NRw4sBLev/view?usp=sharing> (labels) |
| LLaVA-data | <https://huggingface.co/datasets/liuhaotian/LLaVA-Instruct-150K/resolve/main/complex_reasoning_77k.json> (lables)  <https://huggingface.co/datasets/liuhaotian/LLaVA-Instruct-150K/resolve/main/detail_23k.json> (lables)  <https://huggingface.co/datasets/liuhaotian/LLaVA-Instruct-150K/resolve/main/conversation_58k.json> (lables) |
| Objects365 | <https://www.objects365.org/> |
| Private Dataset | Due to the inclusion of unpublished copyrighted content and sensitive information in the dataset, this dataset cannot be presented in the paper. |