

A Appendices

A.1 CLEVR Dataset

Figure 7 shows a topological overview the dataset with sample image, corresponding questions, program chains using the function catalogue.

For detailed information about the function library accompanying the dataset release, please refer to CLEVR VQA (Johnson et al., 2017b). The functions and signatures were kept unaltered from the original specifications⁷.

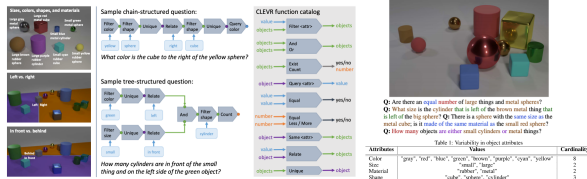


Figure 7: Overview of the CLEVR dataset

A.2 Data

The regular CLEVR dataset can be downloaded from the project page of (Johnson et al., 2017a). We also include a curated dataset ‘1obj’ which contains synthetic data created for a single CLEVR object. In addition to synthetically created questions, images, image scenes, this dataset also contains the images parsed through a semantic image segmentation layer (Mask-RCNN) and a compositional dataset ‘1obj-CoGenT’ which uses distributionally shifted object attributes for training and test data (for the compositional CoGenT test).

A.3 Structural Graph Visualizations from Parser

This section contains additional visualizations for the parser.

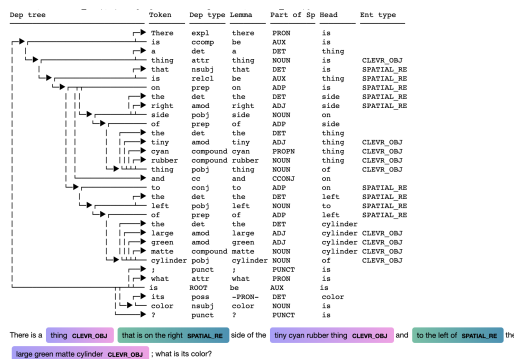


Figure 8: Entity visualization

⁷<https://github.com/facebookresearch/clevr-dataset-gen/blob/master/question-generation/metadata.json>

A.4 Additional Samples

This section contains additional examples from the visualizer.

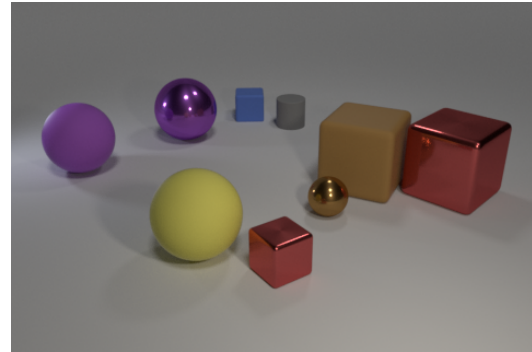


Figure 9: An Image from the CLEVR dataset

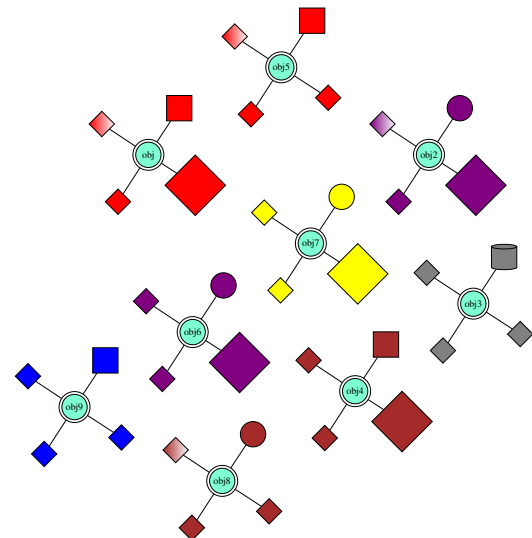


Figure 10: Corresponding structural graph representation of the image

How many spheres in front of the red metal cube and to the right of purple sphere

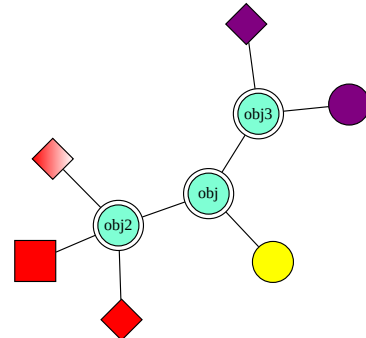


Figure 11: An example question for the corresponding image

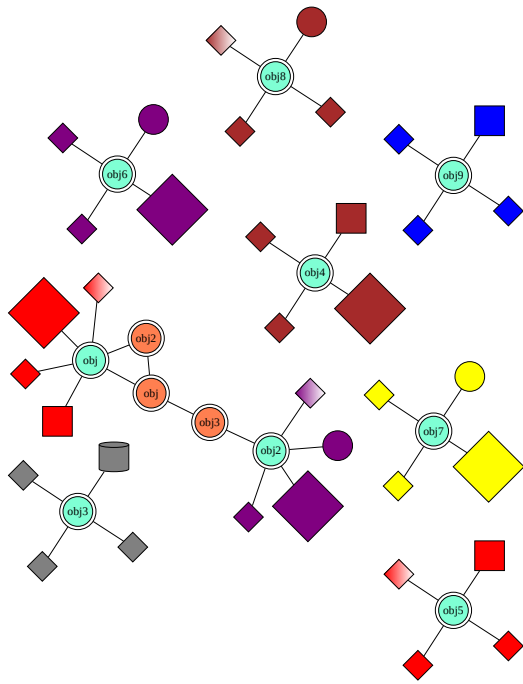


Figure 12: Visualizing both the image and the question in a bipartite graph. Orange nodes represent the text nodes, and the aquamarine nodes represent the image nodes