A Diagnostic Dataset That Contains Minimal Biases

Zhang, Liqiang

May 1,2018

A long-standing goal of artificial intelligence is to project a system which can reason and answer questions about visual information. But answering correctly these questions requires perceptual abilities such as recognizing objects and spatial relationships. However, many show only marginal improvements over strong baselines. In this paper, the author propose a diagnostic dataset for studying the ability of VQA systems to perform visual reasoning. [1]

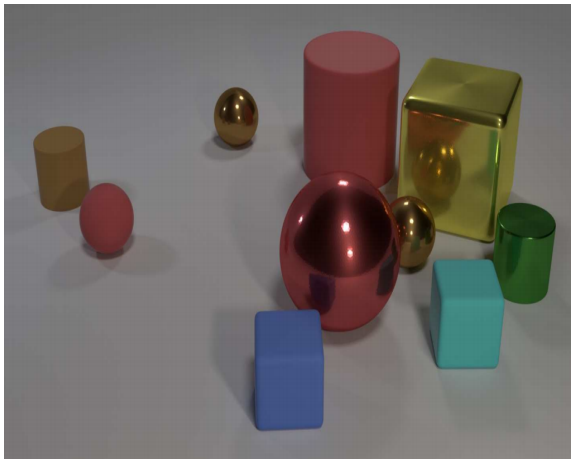

Figure 1: A sample image and questions from CLEVR.

They call it as the Compositional Language and Elementary Visual Reasoning dataset (CLEVR) It has lenging images and questions that test visual reasoning abilities such as counting, comparing, as illustrated in Figure 1. The main components of CLEVR are objects and relationships, scene representation, image generation and question representation and question families and Question generation. CLEVR questions contain two types of relationships: spatial and same-attribute. The author define a question' s size tobe the number of functions in its program and many questions can be correctly answered even when some subtasks are not solved correctly.

The author has introduced CLEVR, a dataset designed to aid in diagnostic evaluation of visual question answering systems by minimizing dataset bias and providing rich ground-truth representations for both images and questions. These experiments demonstrate that CLEVR facilitates in-depth analysis not possible with other VQA datasets. These observations present clear avenues for future work and author plan to use CLEVR to study models with explicit short-term memory.

References

[1] Arijit Ray, Gordon Christie, Mohit Bansal, Dhruv Batra, and Devi Parikh. Question relevance in vqa: Identifying non-visual and false-premise questions. 2016.