Name: Chi Duan
ASUID:1207597985

Generating Synthesis Visual Language Reasoning Data Set

This note book uses a package named ShapeWorld to generate a visual language reasoning reasoning data set. The package ShapeWorld can be accessed from GitHub at the address here. The package generates abstract images consisted by shapes (square, circle, triangle, etc.) in different colors. A (language) caption about the image is also generated. The package can generate images in batches. We first generate the images and then a create manually curated text paragraph, question and answer choices based on each created images. The combined image, paragraph, question and answer choices could be like following.

Image:

Paragraph: If someone add one more circle to the image, there will be two circles in total in the image.

Question: Determine whether the passage is a valid argument for the given image.

Answer: A. True; B. False

Correct Answer: True

Generation Methodology

The manually created formats are used to generate text paragraph, question, answer choices. The formats can be categorized into five main types based on the methods to generate the texts from the image.

1. **Generation by adding one more shape**
    - With this generation method, we add one more quantity of particular shape to the image and generate two types of paragraph. One is correctly describe the the image marked with true after after the addition and the other is falsely describe the image marked with false. We created the paragraph by manually rephrasing the wording in many different ways. The question is to ask whether the paragraph is correct.
2. **Generation by adding one more color**
    - This generation method follows the same way as generation by adding one more shape but in a way to add one more color to the image.
3. **Generation by replacing one shape with another shape**
    - This generation method generates text by assuming that one shape is replaced by another shape in some quantities in the image. The text generation then follows different rephased formats that either give a true or false description of the image after replacing the shape.
4. **Generation by replacing one color with another color**
    - This generation method generates text by assuming that one color is replaced by another color in some quantities in the image. The text generation then follows different rephased formats that either give a true or false description of the image after replacing the shape.
5. **Generation using relational caption**
    - In this method, we use the caption ground truth generated by the ShapeWorld package as a comparing target to generate kinds of true or false statement through various formats.

In order to simplify the dataset generation process due to time limitation, all the answers in this generated data set has a uniform two-option format of ["True", "False"] instead of four-options case.