BITS F312: NEURAL NETWORKS AND FUZZY LOGIC
COURSE PROJECT: VISUAL QUESTION-ANSWERING

Dataset Description:

Link to the training dataset is here!

The dataset provided is a modification of the CLEVR dataset. It consists of a collection of synthesized images and a set of questions associated with each image. Your task is to predict the answers to these questions.

You have been provided with about 15000 images and 135000 questions and answers for training. Most images have about 10 questions associated with them, however, some images don't have any questions associated with them.

The images contain simple 3D objects with each object having one of the 96 property profiles obtained by picking one choice each from the following four *types*:

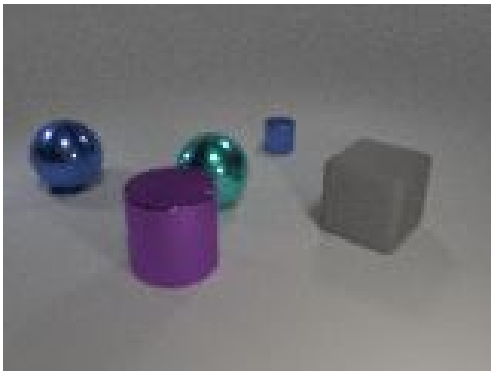
Shape: Sphere, Cube, Cylinder
Size: Large, Small
Material: (Matte) Rubber, (Shiny) Metal
Color: Gray, Cyan, Blue, Purple, Brown, Green, Yellow, Red (8 colors)

The questions test the understanding regarding the various type of relationships between these objects. The relationships include the understanding of the position relative to the other objects i.e which of two objects lies to the left/right of the other or which of the two objects lies in front of/behind the other. Your model should be able to identify the object, based on its relationship with other objects. Your model should also be able to count objects and compare the counts.

The five types of questions asked can be found in Section 4.2 of the paper. Go through Section 3 of the paper for more information on how the images and questions are generated.
The answers to these questions are all one-word answers.

Note: The dataset provided doesn't have any supplementary information (Ex: functional program) other than images and questions (in contrast to the actual dataset).

Example instance from training set:



Question: What color is the large metallic ball that is in front of the metallic ball on the left side of the cylinder in front of the blue sphere?
Answer: **cyan**

Take note of the various baselines mentioned in the paper. This will give you a rough idea of how different models perform.

Readings (might be beneficial):

Knowledge about the following topics might help you improve your model.

1. Attention mechanism

2. Object detection models: YOLO, Fast(er) RCNN

Don't come up complex models right from the offset though, use simple models in the beginning and then keep improving your model. Remember Occam's razor!

Submission:

Deadline: 24th November, 2018 7 PM

You will be required to submit a small write-up on how you ended up with your final model. Make it a point to record your loss plots and accuracies for all your hyperparameter configurations. You'll be able to compare your models and it'll help you submit a sound write-up.

Talk to us about any new ideas you have for your model, we might be able to guide you. You may receive bonus marks for your creativity and thought, provided you are able to justify it (empirically or otherwise).

Submit on this drive link a zip file named as one of the BITS-ID of your team members. It should contain a **maximum of two model files, one submission file and your write-up**.

Instructions for submission.py:

1. Questions are present in the file **Questions.json.** Load the json file. It has two fields *info* and *questions*. Ignore *info. questions* contains a list of dictionary with 3 fields: a) Index b) Question c) Image. Now, the Image field has name of the image file. Actual path of the image is "./images/<name>.png". Question field has the question. Index field has the index for the question.

2. Your **submission.py** file must load the model file from the folder, use the **Questions.json file** to load the question and the image into your model and generate a predicted answer.
3. Finally, it should store the predictions in a **csv file** named **solution.csv** which should have two columns: Index and Answer. Index should contain an *integer* corresponding to the Index of the question. Answer should contain a **lower-case string** representation of the answer predicted.

Desired Format of **solution.csv:**

0,sphere
1,false
2,yellow
3,metal
4,small
5,0 ------> this is a string too.
6,false
.
.
.

Format of **Questions.json**:

{"info": {"license": "Creative Commons Attribution (CC-BY 4.0)", "date": "11/23/2018", "version": "1.0", "split": "new"},
"questions": [{"Index": 0, "Question":
"There is a rubber ball behind the thing that is in front of the small cyan rubber object; what number of cyan spheres are to the left of it?", "Image": "CLEVR_new_000573"},
{"Index": 1, "Question": "Is there a small yellow metal ball behind the sphere behind the tiny cyan rubber object that is to the left of the gray

sphere?", "Image": "CLEVR_new_000573"},
…..]}

Evaluation:

This project accounts for **20%** of your final grade.
The model you submit will be tested and the final marks you obtain will be a function of both **accuracy** and **model size**.

Final Score = 10 * f1(accuracy) + 10 * f2(model_size)

f1(accuracy) = e^(beta*(accuracy - max_accuracy))
f2(model_size) = delta - kappa * model_size

We'll test your model(s) and show you your accuracy score. You can select your final model based on that.